(54) Title: MARKING OF MOVING OBJECTS IN VIDEO STREAMS

(57) Abstract: A method for annotating a video image to include interactive links comprises the step of manually establishing active areas (114) in a subset of the frames of the video image, and using interpolation to establish the active areas for the remaining frames. The active area is an area of a screen (514) where a viewer can place a cursor and click, to thereby activate a link. This link could be a link to a URL or some other kind of link. The interpolation can be a linear interpolation or a spline interpolation technique. The video image can be communicated to a viewing system along with annotation corresponding to some of the video frames but not others of the video frames. The receiving system can determine the active areas for the others of said frames by interpolation.

patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

MARKING OF MOVING OBJECTS IN VIDEO STREAMS

Leon A. Shirman
Yakov Kamen

Background of the Invention

This application pertains to marking of objects in a video stream.

It is known in the art to annotate or "mark" objects in a video stream with an interactive link. For example, suppose one were watching a video program of an automobile race, and there were a number of racing cars shown driving around a track. Also suppose that these cars were marked with interactive links. What this means is that one can position a cursor over the image of one of the cars and click on the image. This will cause the system controlling the video screen to display information that is linked to that image. This link can be in the form of a URL, and the system would display a web page depicting information concerning the car or driver that has been clicked on.

When one "marks" an object, one establishes an area of the video screen surrounding that object (or near that object) as an "active area". If a viewer subsequently clicks on the active area, a link will be invoked. Referring back to the above example, if one of the racing cars is marked, there will be an active area surrounding or near the image of that racing car. If a viewer moves a cursor into that "active area" and clicks on that active area, the above-mentioned link will be invoked.

In a video of a racing car, the car constantly moves, and its position on the video screen constantly changes. Accordingly, the active area corresponding to the car must constantly change, and "keep up" with the car. Thus, each frame of the

1

1   video must be marked with an appropriate active area for that car. Often, this is

2   accomplished by image recognition. In other words, during an editing process, video

3   editing apparatus recognizes if the car has moved from the right side of the video

4   screen to the left side of the video screen. If this has happened, the editing apparatus

5   will establish a new active area for that car on the left side of the screen for those

6   video frames where the car is on the left side of the screen. Unfortunately, there are

7   several problems with image recognition. For example, image recognition systems

8   typically use the color of an object to track its location. If the object moves into an

9   area where the lighting changes, e.g. because a shadow is cast on the object, that can

10 prevent accurate image recognition. In addition, image recognition has other

11 problems that can introduce errors into the process of marking an object.

12   Image recognition is well known in the art and is discussed by J.P. Lewis in

13 "Fast Normalized Cross-Correlation",

14 http://www.idiom.com/~zilla/Papers/nvisionInterface/nip.htm; U.S. Patent 5,267,332,

15 issued to Walch et al.; U.S. Patent 4,611,347, issued to Netravali et al.; U.S. Patent

16 4,868,871, issued to Watson, III; R.O. Duda et al., "Pattern Classification and Scene

17 Analysis", published by John Wiley and Sons in 1973; and R.C. Gonzalez et al.,

18 "Digital Image Processing", 3d edition, published by Addison-Wesley in 1992.

19 Lewis, Walch, Netravali, Watson, Duda and Gonzalez are incorporated herein by

20 reference.

21   Instead of using image recognition apparatus, a person could mark each frame

22 of the video manually. However, marking each frame of the video would be very

23 tedious and time-consuming. What is needed is a novel method for marking moving

24 objects that reduces the amount of human interaction required.

2

1

2   <u>Summary</u>

3   A method in accordance with the invention comprises the step of annotating a

4   video clip with a link. (The video clip can be animation, and as used herein, "video"

5   includes animation.) This link can be a link to a web page or other source of

6   information, or a link to application software. In one embodiment, a first set of

7   frames within a video clip is annotated, e.g. using video editing apparatus. During

8   this step of annotation, a portion of each annotated frame is designated as an "active

9   area." This means that if one moves a cursor to the active area and clicks within the

10   active area, a link associated with the active area will be executed.

11   The active area is typically associated with an image of an object, e.g. a car

12   racing around a track, an actor in a scene, a product displayed in a commercial, a

13   cartoon character, or some other person or thing. Accordingly, during annotation of a

14   frame, a person annotating the video clip will establish the location of the active area

15   associated with that object. (The active area is typically a rectangular or circular

16   region surrounding the object in question.)

17   It would be extremely unwieldy to annotate every frame in a video clip in this

18   manner. Accordingly, in accordance with the invention, only a subset of the frames

19   (e.g. every tenth frame) is annotated. The location of the active area in the other

20   frames is established by interpolation. The location can be established by a linear

21   interpolation technique. Alternatively, the location can be established by an

22   interpolation technique that generates a curve, such as a spline interpolation

23   technique.

24   In another embodiment of a method in accordance with the invention, a subset

25   of frames, for example two frames such as the first and last frames of a sequence of

1   frames, are manually annotated. A video editing system then calculates the position

2   of the active area for each frame between these manually annotated frames of the

3   sequence based on an interpolation algorithm. The system then determines the

4   difference between the interpolated position of the active area and the position of the

5   object with which the active area is associated for each of the frames. This

6   calculation is based on image recognition. The image recognition can be carried out

7   by image recognition software. Alternatively, image recognition can be performed by

8   a combination of hardware and software. (Image recognition can also be performed

9   by hardware. The partitioning between hardware and software can be done in

10   different ways.) Thereafter, the system highlights those frames that the system has

11   determined to exhibit the greatest error. In other words, the system highlights those

12   frames in which the distance between the active area and the object associated with

13   that active area is greatest. One skilled in the art can define a distance between an

14   active area and the object based on maximum, minimum or average distance between

15   selected area and points within the object.

16   In one embodiment, the system flags the frame with the largest error. In

17   another embodiment, the system flags the n frames with the n largest errors, where n

18   is an integer. In another embodiment, the system flags all frames in which the error is

19   over a certain threshold.

20   Thereafter, the person annotating the video clip can observe the flagged

21   frames and decide whether he or she desires to manually annotate those frames. The

22   person annotating the video clip also has the option of recalculating the interpolation.

23   After recalculating the interpolation, the user can see if other frames are flagged that

24   might require manual annotation.

4

1   Flagging may be accomplished by highlighting the flagged frame on a display

2   or by providing some other form of output indicator.

3   In accordance with another aspect of the invention, a video signal is sent to

4   one or more video viewing systems so that one or more viewers can watch the video

5   signal. The video signal can be broadcast to many viewers. The location of one or

6   more active areas associated with a video signal is also communicated to the one or

7   more video viewing systems for some of the frames of the video signal, but not other

8   frames of the video signal. The one or more video viewing systems determine the

9   location of the one or more active areas for the other frames by interpolation. In this

10   way, instead of having to provide data to the video viewing systems concerning the

11   active area locations for each frame, such data is provided only for selected frames,

12   and the video viewing systems calculate where the active areas for the other frames

13   should be by interpolating, e.g. using either a linear interpolation technique or a spline

14   interpolation technique. Because the source of the video signal need not provide

15   active area information for each frame of the signal, the bandwidth requirements for

16   communicating this information are reduced.

17   In one embodiment, not only is the location of the active area communicated

18   in this way, but the size of the active area is communicated in this way as well.

19   In accordance with another aspect of the invention, a video signal comprising

20   a set of frames is communicated along with the location of an active area for some of

21   the frames but not others of the frames. The active area corresponds to an object

22   depicted in those frames. The object can be a product, a person, a cartoon character,

23   or anything else that can be depicted in a video. As mentioned above, the active area

24   is not communicated for the others of the frames, although the object is depicted in

25   the others of the frames. A receiving system calculates the location of the active area

5

1   by interpolation based on the location of the active area for said some of the frames.

2   In this way, the bandwidth requirements for communicating the location of the active

3   areas is reduced.

4

5   Brief Description of the Drawings

6   Fig. 1 illustrates a display screen of a video editor/player used to annotate a

7   video clip in accordance with the present invention.

8   Fig. 2 illustrates the position of an object on a video screen along the y

9   direction vs. frame number for a sequence of frames in a video clip.

10   Fig. 3 is a flow diagram illustrating a method in accordance with a first

11   embodiment of our invention.

12   Fig. 4 illustrates the data architecture for a video editor/player used to mark a

13   video clip in accordance with the invention.

14   Fig. 5 is a block diagram of a signal transmission system and a reception and

15   display system in accordance with another embodiment of the invention.

16   Fig. 6 illustrates the position of the active area along the y axis vs. the frame

17   number of a video clip.

18

19   Detailed Description

20   Fig. 1 shows a display screen 100 depicting a typical graphical user interface

21   ("GUI") for marking editors. Such an interface can be used during a method in

22   accordance with our invention. Screen 100 shows a sequence of so-called key frames

23   101a to 101p in the editing area. As is known in the art, video images comprise a

24   sequence of frames. For digital video, each frame corresponds to an array of pixels.

25   Each pixel is a data value corresponding to the color, brightness and saturation of a

6

1   small area of the frame. In certain video compression or encryption schemes, instead

2   of storing every pixel array of every frame, the entire pixel array is stored only for

3   certain "key" frames. Data is stored for the other frames that merely represent the

4   difference between the image in those other frames and the preceding key frame. An

5   example of a compression technique that uses this scheme is MPEG-2 (motion

6   pictures expert group standard 2).

7   For the case of linearly digitized analog video that is not compressed or

8   encrypted, every nth frame (e.g. every 10th frame) can be chosen as a key frame

9   (along with frames corresponding to scene changes). As mentioned above, screen 100

10   depicts a sequence of key frames 101a to 101p.

11   Arrows 104a to 104d are not depicted on screen 100. Rather, arrows 104a to

12   104d are included in Fig. 1 merely to symbolize the chronological sequence of the

13   frames.

14   Screen 100 includes a region 106 that depicts one of key frames 101a to 101p.

15   The image in region 106 is magnified compared to the depiction of key frames 101a

16   to 101p elsewhere on the screen, and represents the frame that is currently being

17   edited. In the example of Fig. 1, key frame 101g is being edited. This is indicated by

18   the presence of an icon or mark 101g' provided at frame 101g.

19   Typically, screen 100 also depicts navigation bars and toolbars for functions

20   such as fast forward, fast backwards, new file, etc. The system driving display 100

21   includes software drivers for supporting the navigation bars and toolbars. These

22   navigation bars and toolbars are not shown in Fig. 1 for sake of clarity.

23   Also shown on screen 100 is a menu box 108 which enables a user to select a

24   shape of an active area about an object that a viewer will be able to click on. Box 108

25   enables an editor to select among many different shapes for the active area, only some

7

1   of which are depicted in Fig. 1. As mentioned above, the active area is the location on

2   a screen where an object of interest (e.g. a car or a person) is displayed. By clicking

3   within the active area associated with that object or person, one effectively clicks on

4   that object or person.

5   (When a viewer ultimately watches the video clip, the active area may or may

6   not be displayed. For example, the system on which the video is displayed may

7   highlight the active area. Alternatively, other systems on which the video is displayed

8   may not highlight the active area. In yet other systems, the viewer may have the

9   option to decide whether or not to highlight the active area.)

10   Referring back to Fig. 1, box 110 is a menu box that allows the editor to

11   introduce the link into the image, which is typically a URL. However, other types of

12   links can be provided, including specialized links for launching desired application

13   software.

14   As seen in Fig. 1, region 106 depicts a person 112, whose head 113 has been

15   marked by a square 114 and linked to a link L1. (The square shape has been selected

16   in menu box 108, and link L1 has been selected in menu box 110.) Square 114

17   symbolizes the active area. During a scene in a video, objects (e.g. the head of a

18   person) typically move in a limited, non-jerked manner. (Good camera control during

19   filming of a movie, television show or other video typically involves minimizing any

20   jerking of the object of interest on the screen. Also, a small object moving in a jerked

21   fashion, even if closely followed, would be difficult for a viewer to target when

22   viewing the annotated video.)

23   Square 114 represents the active area of the video screen on which a viewer

24   will click if he wants to access information linked to person 112. For example, if the

25   video is a soap opera, and the viewer wants to obtain more information about the

8

1   character played by person 112, he will click on person 112, and in particular, he will

2   position the cursor within the active area represented by square 114

3   During the video, person 112 typically moves about the set. Accordingly, it is

4   necessary to have active area 114 move about the screen. This is necessary because if

5   person 112 walks from the left side of the video screen to the right side of the video

6   screen, and the viewer wants to find out more about person 112, he is going to want to

7   click on current location of person 112.

8   In principal, one could annotate each frame of the video to determine where

9   active area 114 should be positioned, but this would be an enormously tedious, time-

10   consuming task. Accordingly, it would be desirable to simplify this task. In

11   accordance with the present invention, some of the frames are annotated by placing

12   active area 114 at an appropriate location. (This can be done manually or through

13   other techniques.) The other frames are annotated by an interpolation method

14   described below.

15   Fig. 2 illustrates the position of head 113 in a video along the y direction for

16   each frame in a series of frames from a start frame A to an end frame D. Point 203

17   represents the starting position of head 113 at the start of a scene and point 204

18   represents the ending position of head 113 at the end of the scene. Line 206 traces the

19   position of head 113 between points 203 and 204 as the scene progresses. (Head 113

20   changes position, e.g. as person 112 moves about the set during the video.)

21   Line 208 shows a direct linear interpolation, resulting by having a straight line

22   between the two positions 203 and 204. In other words, rather than annotating each

23   frame manually, the position of active area 114 is automatically inserted into each

24   frame by performing linear interpolation between positions 203 and 204. As can be

25   seen, this is not necessarily the most accurate way of annotating the frames. One can

9

1   improve the linear interpolation by manually annotating more frames, and linearly

2   interpolating between each frame for which a manual annotation has been performed.

3   Fig. 2 also shows intermediate points 216, 218, indicating the position of head

4   113 in frames B and C, respectively. In one embodiment, the position of active area

5   114 is manually inserted (e.g. visually) for a set of frames, e.g. frames A, B, C and D.

6   The position of active area 114 is established by linear interpolation between each

7   point for which the position of active area 114 has been manually annotated. Thus, in

8   Fig. 2, lines 224, 226 and 228 are the result of linearly interpolating between points

9   203 and 216, points 216 and 218, and points 218 and 204 respectively.

10   A more accurate way of interpolating is to use an interpolation technique that

11   results in the generation of a curve that represents the position of the object from

12   frame to frame. An example of such a technique is spline interpolation. When

13   performing spline interpolation, one typically only needs the end points and one or

14   two intermediate points, and the spline interpolation technique will accurately track

15   the object. The spline will smooth the motion and eliminate jitter often seen by image

16   recognition.

17   It will be appreciated that one must not only establish the position of an object

18   in the y direction, but also the x direction. Thus, during a method in accordance with

19   our invention, the position of the active area will be established by interpolation for

20   both the y and x directions. In addition, the object associated with the active area can

21   expand or contract during the course of a video scene, e.g. because it moves closer to

22   or further from the camera. Accordingly, in one embodiment, the size of active area

23   114 is also manually annotated for several frames during a scene, and the size of the

24   active area is thereafter interpolated in the same manner as the position of the object.

10

1   Fig. 3 is a block diagram illustrating a method in accordance with a first

2   embodiment of the invention. Referring to Fig. 3, the method begins by loading a

3   block of video frames into a memory (step 301). The memory can be a magnetic

4   memory (e.g. a magnetic disk, magnetic tape or magneto-optic disk). The memory

5   can also be an optical disk or a semiconductor memory such as a DRAM, SRAM,

6   EPROM, or other type of memory. The video information can be either in analog

7   form or in digital form.

8   Thereafter, one manually annotates a subset of those frames (step 302).

9   During this step of manual annotation, a set of frames are typically displayed on

10   portions 101a to 101p of the screen 100, and one of those frames can be selected for

11   annotation at a time. The selected frame is shown on portion 106 of screen 100. The

12   location, size and shape of the active area are established within those frames that are

13   being annotated, and a link associated with that active area is established. Manual

14   annotation can be performed using an input device such as a mouse, touchpad,

15   joystick, trackball, cursor control keys, or other input device of the video editing

16   system. During this step, the annotation information is stored in a memory, e.g. as

17   described below.

18   For the frames that are not manually annotated, annotation is done by

19   interpolation (step 303). This can be done by any of the following steps:

20   a)   linear interpolation of the position of the active area and linear

21   interpolation of the size of the active area;

22   b)   linear interpolation of the position of the active area with no change in

23   size of the active area;

24   c)   spline interpolation of the position of the active area and spline

25   interpolation of the size of the active area; or

11

1   d)   spline interpolation of the position of the active area with no change in

2   the size of the active area.

3   In lieu of linear or spline interpolation, other interpolation techniques can be used. By

4   way of example, interpolation can be:

5   a)   between the start and end frames in a video clip; or

6   b)   interpolation from each manually annotated frame to the next manually

7   annotated frame. (Typically, the start and end frames of the clip are

8   among the frames that are manually annotated.)

9   Referring to block 304, it is possible that interpolation can produce an error in

10   position of the active region. In other words, there will be a difference between the

11   position of an object and the position of the active area associated with that object.

12   After interpolation, a person annotating the video clip can view the location of the

13   active areas for the various frames at portions 101a to 101p of screen 100. If one of

14   the frames shows a large discrepancy between the position of object 113 and active

15   area 114, the person editing the video clip can manually annotate that frame, and have

16   the editing system re-compute (e.g. by interpolation) all of the active area positions

17   for the frames that have were not manually annotated.

18   After establishing the position of active area 114, the person annotating the

19   video clip chooses the form of the links represented for users. Instead of using a box,

20   circle, ellipse or other shape, the active area could by depicted as a logo (e.g. a GIF

21   file), a piece of animation (e.g. an animated GIF file) or other video stream segment.

22   (GIF is an acronym for graphics interchange format.) As mentioned above,

23   optionally, the active area can be visually indicated when the annotated video clip is

24   later viewed.

12

1   Referring to block 306, either after or before interpolation, a link to a URL can

2   be associated with each active area. Thus, when a viewer clicks on that active area,

3   that URL will be invoked. As mentioned above, the link need not be to a URL. It can

4   also be a link to application software, or a source of information stored at a particular

5   address.

6   Fig. 4 illustrates a data architecture for an embodiment of a video editor/player

7   used to mark a video clip. Such a video editor/player comprises a data mass storage

8   device 406 that contains numerous video clips. The data mass storage device can be a

9   magnetic disk, a magneto-optic disk, an optical disk, a semiconductor memory, or

10   other type of memory storage device.

11   Only one video clip 420a is shown in Fig. 4 for clarity, along with one

12   matching annotation file 421a, although more than one video clip and annotation file

13   may be stored in memory device 406. There can be more than one annotation file per

14   video clip. The annotation file contains such things as the spline (either as a formula

15   or as a reference to a predefined set of spline types) and its associated parameters, the

16   link itself, and the color and size of the marker of the active area. Element 450 is a

17   table or database, containing all the associated labels for each video sequence of

18   frames, thus serving as a directory for the video presentation. There may be, of

19   course, several databases of shows and annotation files, possibly organized in a

20   hierarchy structure, and they may share some of the files 420 or 421 depending on the

21   edits done.

22   In one embodiment, during annotation, data is loaded into a memory

23   corresponding to the frame number, the location of the active area, its shape, its size,

24   and the URL or other link associated with that active area. That data is provided to

25   the viewer along with the video signal, typically in a separate file. This information

13

SUBSTITUTE SHEET (RULE 26)

1   can be provided as a separate data stream to the viewer. The viewer's video display

2   system tracks the frame number, and for each frame, determines whether a cursor is

3   positioned within one of the active areas established for that frame.

4   In another embodiment, information concerning the location of the active area,

5   its shape, its size, and the link are communicated and embedded as part of the video

6   signal. This information can be provided during one of the blanking intervals of the

7   signal or during some other portion of the signal that is not typically used.

8

9   <u>Embodiment In Which Interpolation Errors Are Minimized or Corrected</u>

10   As can be seen in Fig. 2, when using interpolation to establish an active area

11   on a video screen, there can be errors between the position of the marked object and

12   the position of the active area. In accordance with another embodiment of the

13   invention, these errors can be reduced or minimized through the following technique.

14   1.   A block of video frames is loaded into a memory (step 301 of Fig. 3).

15   2.   A first set of n frames is manually annotated, where n is greater than or

16   equal to 2 (step 302 of Fig. 3). The n frames can be the first and last frames in a video

17   sequence.

18   3.   Interpolation is performed to establish an estimated position of an

19   active area for those frames within the video sequence that are not manually annotated

20   (step 303 of Fig. 3). Interpolation can be by linear interpolation. Alternatively,

21   interpolation can be by a technique that results in a curve, e.g. a spline technique. As

22   mentioned above, several interpolation techniques can be used to generate lines or

23   curves, and the person annotating the video clip can choose among them.

24   4.   For each frame that is not manually annotated, the video system

25   calculates the distance between the active area and the object that the active area is

1   associated with. (The position of the active area is established by interpolation. The

2   position of the object is established by automated image recognition techniques, e.g.

3   as discussed above. As indicated above, image recognition may or may not be

4   accurate.)

5   5.   The system flags those frames that have the greatest distance between

6   active area 114 and the object 113 that the active area is associated with. The system

7   can highlight those frames (e.g. one or more frames 101a to 101p) that exhibit these

8   distances and can place an icon or other marker at those frames, or it can use some

9   other technique for identifying these frames to the user. The user can look at these

10   frames, and verify whether the active area is, in fact, an unacceptable distance from

11   the object with which it is associated.

12   6.   If the user believes that this distance is unacceptable, he or she can

13   manually annotate the frames exhibiting the unacceptable error (step 304 of Fig. 3).

14   7.   After manually annotating the frames exhibiting the unacceptable

15   error, the user can cause the system to recompute (e.g. by again interpolating) the

16   position of the active area based on data including the new manual annotation.

17   8.   Optionally, the user can have the system take the new interpolated

18   active area positions and establish another list of frames having the largest errors. If

19   the user finds these errors to be acceptable, or merely artifacts relating to image

20   recognition errors, he or she can ignore them.

21   9.   One then associates an active area shape (and optionally a visual

22   indicator) and a link with the active area.

23   As mentioned above, during step 304 of Fig. 3, one can cause the system to

24   compute the distance between the active area and the object associated with that

25   active area, and flag those frames shown at portions 101a to 101p of screen 100

SUBSTITUTE SHEET (RULE 26)

1   having unacceptable errors. The user can manually annotate those frames exhibiting

2   unacceptable errors as described above. (This process can be used to adjust both the

3   position (along both the x and y axes) and size of the active areas.)

4

5   <u>Communicating Information Regarding the Active Areas to a Television Viewer</u>

6   After the video frames are marked with information concerning the active

7   areas, it is necessary to communicate that information to the viewer's video system.

8   Fig. 5 illustrates a video transmission system 500 comprising a data mass storage

9   device 506 including a portion 520 that stores a video program to be transmitted. The

10   video program can be any type of program—sporting events, soap operas, educational

11   programs, cartoons and other animation, etc. Also included within mass storage

12   device 506 is a portion 521 that contains annotation information. This annotation

13   information includes a) the location of several active areas associated with various

14   frames of the video program; b) the size of those active areas; and c) the link that

15   those active areas are associated with. (The link can be in the form of a URL.) In

16   other words, the annotation information might indicate that in a particular frame of the

17   video program, there is a first active area located in the upper left portion of the video

18   frame, and that active area has a certain size. If the viewer clicks on that active area,

19   that should activate a link to a particular web page.

20   Mass storage device 506 can be any type of mass storage device, e.g. a

21   magnetic disk, magnetic tape, optical or magneto-optic media, or other data storage

22   structure.

23   Information from mass storage device 506 is provided to a data separation

24   circuit 505 that separates out the video signal and provides that video signal to a

25   conventional transmitter 504. (Data separation circuit 505 is typically a server.) A

16

1   television system 501 includes a receiver 511 for receiving the video signal. While

2   Fig. 5 includes symbols indicating that the video signal is transmitted and received as

3   conventional television radio signals, the video signals can be communicated by other

4   means, e.g. an electrical cable, optical cable, microwave satellite communication,

5   telephone lines, the internet, or other signal communication technique.

6   The annotation information (e.g. the location and size of the active area, along

7   with the URL information associated with the active area) is communicated to

8   televisions system 501 via a communication medium 503, which can be the same as,

9   or different from, the medium by which the video signal is communicated. Thus,

10   communication medium 503 can be conventional radio signals, an electrical cable,

11   optical cable, microwave, satellite communication, telephone line, the internet, or

12   other signal communication technique.

13   The annotation information is received by a preprocessing unit 512.

14   Preprocessing unit 512 calculates, via an interpolation technique, the location of each

15   active area of each frame for those frames in which that data was not provided by

16   transmission system 500. In one embodiment, the interpolation technique is a linear

17   interpolation technique, e.g. as described above with respect to which Fig. 2. For

18   example, suppose transmission system 500 provides the location for an active area

19   every tenth frame. Fig. 6 illustrates that the y coordinate values for the active area for

20   frames 0, 10, 20 and 30, indicated by points 600, 601, 602 and 603, respectively. If a

21   linear interpolation technique is used, it will be assumed that the y coordinate for the

22   active area will be as established by lines 604, 605 and 606. Other interpolation

23   techniques can be used, e.g. techniques that generate a curve instead of a straight line

24   between specific points. For example, a spline interpolation technique can be used. If

25   a spline interpolation technique is used, depending upon the spline calculation

17

SUBSTITUTE SHEET (RULE 26)

1   parameters, it will be assumed that the y coordinate for the active area will be along

2   curve 607.

3   Fig. 6 illustrates the y coordinate of the active area vs. frame number. Similar

4   calculations are performed for the x coordinate of the active area, and the size of the

5   active area. (The size of the active area can increase or decrease, e.g. depending on

6   whether the object associated with the active area moves toward or away from the

7   camera.)

8   Referring again to Fig. 5, it is seen that the preprocessing unit 512 provides the

9   video signal to a video display unit 514. Preprocessing unit 512 is coupled to a user

10 interface. The user interface 513 permits a user to control certain features of the

11 image displayed on display unit 514, e.g. whether a marker will be shown on the

12 screen indicating the location of the active area. User interface 513 also permits a

13 user to click on the various active areas to activate the links associated therewith.

14   User interface 513 typically has an appropriate control mechanism such as a

15 trackball, mouse, touchpad or joystick. This interface is used to position a visible

16 cursor on video display unit 514 so that the viewer knows what object he is clicking

17 on. Preprocessing unit 512 determines whether the cursor has been positioned over an

18 active area. In one embodiment, preprocessing unit 512 comprises a microprocessor

19 system.

20   As mentioned above, in one embodiment of the invention, the active areas for

21 selected frames are established, e.g. using a manual annotation system. The active

22 areas are associated with objects depicted in the video frames. During this method,

23 the location, size and shape of each active area is established for some (but not all) of

24 the frames of the video clip. For other frames of the video clip, the location and size

25 of the active areas is established by interpolating between the frames. In one

18

1   embodiment, the location and size of the active areas for those frames that were

2   manually annotated are communicated to video system 501. In other embodiments,

3   the frames for which active locations and sizes are communicated are selected on

4   other bases. For example, the location and size of the active areas can be

5   communicated to receiving video system 501 for every tenth frame. Alternatively, the

6   location and size of the active areas can be communicated only for the so-called "key

7   frames" of the video clip. (As mentioned above, certain video compression

8   algorithms communicate the entire pixel array only for certain key frames. For other

9   frames, only information corresponding to the change between one frame and the next

10   is communicated. The key frames are those frames for which the entire pixel array is

11   communicated.)

12   Referring again to Fig. 5, in one embodiment, a workstation 550 can be

13   coupled to video transmission system 500. Workstation 550 comprises a personal

14   computer ("PC") 551 and a local memory storage device 552 (e.g. a disk) or other

15   non-local storage (e.g. a server) that contains editing software 553. Workstation 550

16   is used to annotate the video clip. In this example, workstation 550 is coupled via a

17   communication line 554 to server 505, but many other ways to network workstation

18   550 to server 505 can be used. In some embodiments, editor software 553 can run on

19   server 505. In other embodiments, the files (both the video and annotation files) may

20   reside on the local memory 552.

21   While the invention has been described with respect to specific embodiments,

22   those skilled in the art will appreciate that changes can be made in form and detail

23   without departing from the spirit and scope of the invention. For example, the pixels

24   of the video frames can be stored in terms of RGB values. (RGB refers to the values

25   of the red, blue and green color components associated with the pixel.) Alternatively,

19

1   the pixel values can be stored in terms of color, saturation and intensity. Screen 100

2   and/or video display unit 514 can be a CRT screen, and LCD, a projection display, or

3   other appropriate display technology. Different types of video editing/annotating

4   systems can be used to practice the method of the present invention. The

5   interpolation can be performed by a microprocessor or other computing device within

6   the system.

7   The method of the present invention can be used to provide or receive more

8   than one active area in a video clip.

9   Frames can be chosen for manual interpolation in different manners. For

10 example, in one embodiment, every nth frame is manually interpolated, where n is an

11 integer (e.g. 10).

12   The apparatus chosen to perform the interpolation and editing can be a

13 microprocessor-based system with a display screen executing video editing and/or

14 image recognition software programs.

15   In one embodiment, the first and last frames are not manually annotated.

16 Accordingly, in such embodiments, the location and size of the active areas are

17 extrapolated.

18   When the video signal is sent to a viewer, both the video and annotation

19 information can be sent as separate signals to receiving system 501, or they can be

20 communicated as parts of a combined signal. The combined signal can be provided

21 by transmitter 504 via broadcast medium 502. Alternatively, the combined signal can

22 be provided via any of the above-mentioned communications media, e.g. optical,

23 cable, microwave transmission, satellite transmission, etc.

24   The various aspects of the invention can be practiced independently of one

25 another. For example, the annotation apparatus of Fig. 1 can be used without using

SUBSTITUTE SHEET (RULE 26)

1   the signal communication apparatus of Fig. 5. Similarly, the communication

2   apparatus of Fig. 5 can be used to communicate video clips annotated using methods

3   other than the method described above with respect to Figs. 1 to 4. Accordingly, all

4   such changes come within the present invention.

21

1   We claim:

2   1.   A method for establishing the location of an active area within a video

3 clip, said video clip comprising a set of frames, said method comprising the acts of:

4   (a)   establishing the location of said active area for a first plurality of

5 frames within said set of frames;

6   (b)   establishing the location of said active area for a second plurality of

7 said frames within said set of frames by interpolation based on the location

8 established during act (a) above.

9

10   2.   Method of claim 1 wherein the locations established during the acts (a)

11 and (b) are stored in a memory device.

12

13   3.   Method of claim 1 wherein act (a) comprises:

14   displaying said first plurality of frames on a display device; and

15   manually annotating said frames by storing within a memory information

16 concerning the location of said active area within said first plurality of frames.

17

18   4.   Method of claim 1 wherein said interpolation is linear interpolation.

19

20   5.   Method of claim 1 wherein said interpolation is spline interpolation.

21

22   6.   Method of claim 1 wherein said first plurality of frames are key

23 frames.

24

22

1   7.   Method of claim 1 wherein said interpolation is performed by a

2   computing device within an editing system.

3

4   8.   Method of claim 1 further comprising the acts of:

5   (c)   displaying the active area for a third plurality of frames within said set

6   of frames;

7   (d)   annotating at least some of said frames within said second plurality;

8   and

9   (e)   establishing the position of said active area within said second plurality

10   but not within said some of said frames by interpolation after performing act (d).

11

12   9.   Apparatus comprising:

13   a memory device storing a video clip, said video clip comprising a plurality of

14   video frames;

15   an input device for inputting information into said apparatus indicating the

16   location of an active area within some of said video frames; and

17   a computing device within said apparatus for establishing the location of said

18   active area for other frames within said plurality.

19

20   10.   Apparatus of claim 10 wherein said computing device establishes the

21   location of said active area for said other frames within said plurality of interpolation.

22

23   11.   A method comprising:

23

1   providing a video clip comprising a set of frames, a first plurality of frames

2   within said set being annotated with information indicating the location of an active

3   area, said active area being associated with an object shown within said video clip;

4   establishing location data for said active area for those frames that are not

5   within said first plurality by interpolation;

6   using automated image recognition to identify one or more of said frames that

7   are not within said first plurality for which the distance between said active area and

8   said object is large compared to others of said frames that are not within said first

9   plurality.

10

11   12.   Method of claim 11 wherein said automated image recognition identifies

12   frames in which the distance between said active area and said object exceeds a

13   threshold value.

14

15   13.   Method of claim 11 wherein said automated image recognition identifies the n

16   frames for which the distance between the active area and object are largest, where n

17   is an integer.

18

19   14.   Method of claim 11 further comprising displaying said one or more frames on

20   a display so that a viewer can see the distance between said active area said object in

21   said one or more frames.

22

23   15.   Method of claim 11 further comprising annotating at least one of said one or

24   more frames after said act of using.

25

24

1   16.   Method of claim 15 further comprising establishing location data for said

2   active area for those frames that are not within said first plurality and are not within

3   said at least one frame by interpolation.

4

5   17.   Method of claim 16 further comprising displaying the frames whose location

6   data has been established by interpolation so that a user can see how far the active

7   area is from the object.

8

9   18.   Method of claim 11 wherein an operator annotates said first plurality of frames

10   with said information prior to said act of establishing.

11

12   19.   Method of claim 18 wherein said annotation is performed by storing data in a

13   memory indicating the position of said active area for said first plurality of frames.

14

15   20.   Method of claim 11 wherein said video clip is stored in a memory device.

16

17   21   Method of claim 20 wherein said information indicating the location of an

18   active area is stored in the same memory device as said video clip.

19

20   22.   Method of claim 21 wherein said video clip and said information indicating

21   the location of an active area are located in different files within said memory device.

22

23   23.   Method of claim 21 wherein said information indicating the location of said

24   active area is embedded within said video clip.

25

25

SUBSTITUTE SHEET (RULE 26)

1   24.   Method comprising the steps of:

2   providing a video clip, said video clip comprising a group of frames, said

3   video clip depicting an object;

4   annotating some of said frames within said group of frames, but not others of

5   said frames within said group of frames, by storing in a memory the location of an

6   active area within said some of said frames associated with said object;

7   establishing the location of said active area for said others of said frames by

8   interpolation;

9   using automated image recognition to determine those frames within said

10 others of said frames for which the distance between said active area and said object

11 is large compared to others of said frames that are not within said first plurality.

12

13 25.   Apparatus comprising:

14   a memory structure for storing a video clip, said video clip comprising a set of

15 frames, a first plurality of frames within said set being annotated with information

16 indicating the location of an active area, said active area being associated with an

17 object shown within said video clip;

18   means for establishing location data for said active area for those frames that

19 are not in said first plurality by interpolation;

20   means for using automated image recognition to identify one or more of said

21 frames for which the distance between said active area and said object is large

22 compared to others of said frames that are not within said first plurality.

23

24 26.   A method comprising the steps of:

25   receiving a video image;

26

1   displaying said video image as a sequence of frames;

2   receiving information corresponding to the location of at least one active area

3   for some of said frames but not others of said frames; and

4   determining the location of said active area for said others of said frames

5   based on said received information.

6

7   27.   Method of claim 26 wherein said step of determining is performed by linear

8   interpolation.

9

10   28.   Method of claim 26 wherein said step of determining is performed by spline

11   interpolation.

12

13   29.   Method of claim 26 wherein information corresponding to the size of said

14   active area is received for said some of said frames but not said others of said frames,

15   said method also comprising the step of determining the size of said active area for

16   said others of said frames by interpolation.

17

18   30.   Method of claim 26 wherein said active area is associated with a link, said

19   method further comprising the step of activating said link.

20

21   31.   Method of claim 26 wherein said link is a link to an internet page.

22

23   32.   Method of claim 26 further comprising activating said link.

24

SUBSTITUTE SHEET (RULE 26)

1   33.   Method of claim 32 wherein said displaying comprises displaying said video

2   on a display screen and said activating comprises clicking on an active area on said

3   display screen.

4

5   34.   Method comprising:

6   communicating a video clip comprising a set of video frames, some of the

7   video frames within said set of video frames being annotated such that an active area

8   is established for said some of said video frames;

9   communicating the location of said active area for said some of the video

10   frames within said set of video frames and not others of said video frames within said

11   set of video frames.

12

13   35.   Method of claim 34 wherein said video clip is communicated through a first

14   communication medium and said location of said active area is communicated

15   through a second communication medium that is different from said first

16   communication medium.

17

18   36.   Method of claim 34 further comprising:

19   receiving said video clip and the location of said active area for some of said

20   video frames within said set of video frames and not said others of said video frames;

21   and

22   establishing positions for said active area within said others of said video

23   frames by interpolation.

24

28

SUBSTITUTE SHEET (RULE 26)

1   37.   Method of claim 36 wherein said active area is associated with an object, said

2   object being shown in each of said video frames within said set of video frames.

3

4   38.   A system comprising:

5   a receiver for receiving a video signal;

6   a display coupled to said receiver for displaying a video image corresponding

7   to said video signal, said video image comprising a set of frames;

8   a circuit receiving information indicating the position of at least one active

9   area for some of said frames and not others of said frames, said circuit determining

10   the location of said active area for said others of said frames by interpolation based on

11   the location of said active area for said some of said frames.

12

13   39.   System of claim 38 further comprising a transmission system for transmitting

14   said video signal and said information corresponding to the position of at least one

15   active area for said one of said frames and not said others of said frames.

16

17   40.   Apparatus comprising:

18   a first memory storage device storing a video signal, said video signal

19   comprising a set of frames;

20   a memory storage device storing the location of an active area for some of said

21   frames;

22   a transmitter for transmitting said video signal and said location of said active

23   area for said some of said frames but not others of said frames, whereby a system

24   receiving said video signal and said location of said active area for said some of said

SUBSTITUTE SHEET (RULE 26)

1   frames can establish a location for said active area for said others of said frames by

2   interpolation.

3

4   41.   Apparatus comprising:

5   a memory storing a video clip;

6   a structure for establishing the location of an active area for some of the

7   frames of the video clip but not other frames of the video clip; and

8   a transmission system for transmitting the video clip and the location of the

9   active area for said some of said frames of the video clip but not other frames of the

10   video clip.

11

12   42.   Apparatus of claim 41 further comprising a receiving unit for receiving and

13   displaying the video clip and establishing the location of said active area for said other

14   frames by interpolation.

15

16   43.   A method for establishing the location of an active area within a series of

17   frames of images, said method comprising the acts of:

18   (a)   establishing the location and/or size of said active area for a first

19   plurality of frames within said series of frames;

20   (b)   establishing the location and/or size of said active area for a second

21   plurality of said frames within said series of frames by calculation based on the
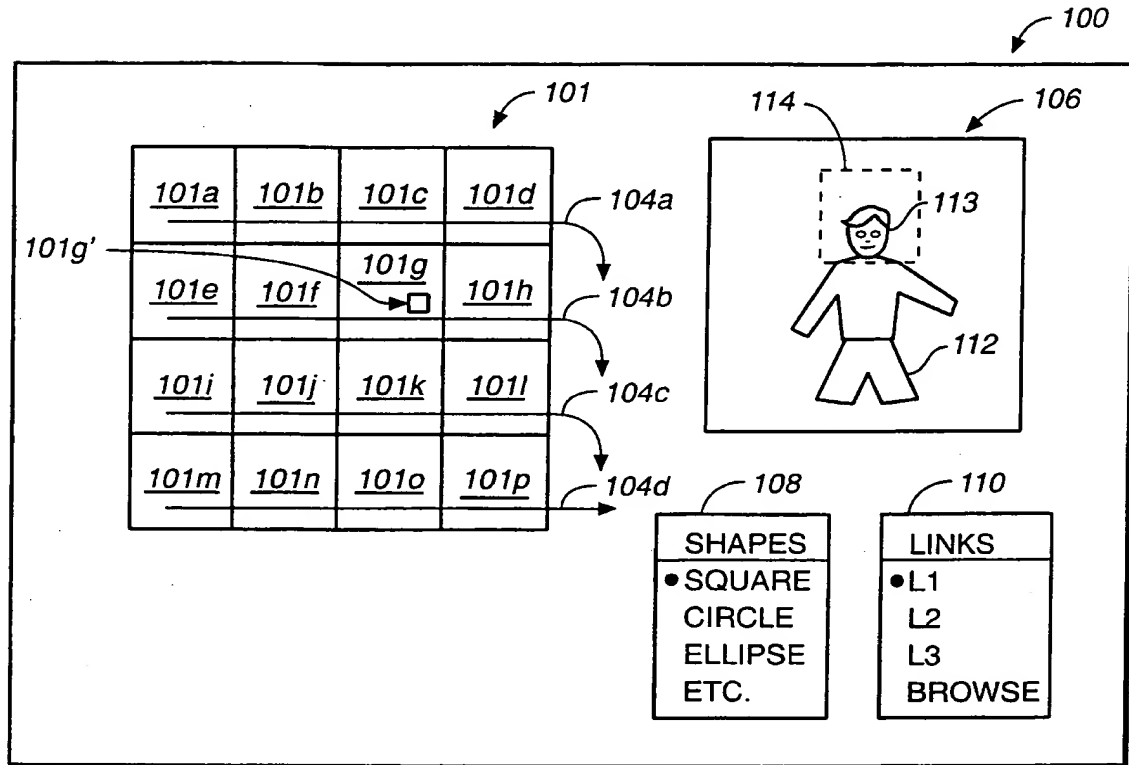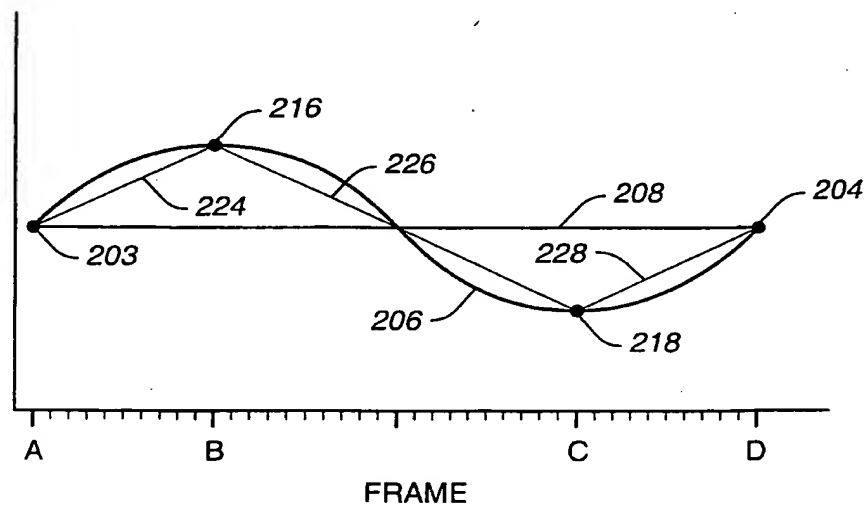
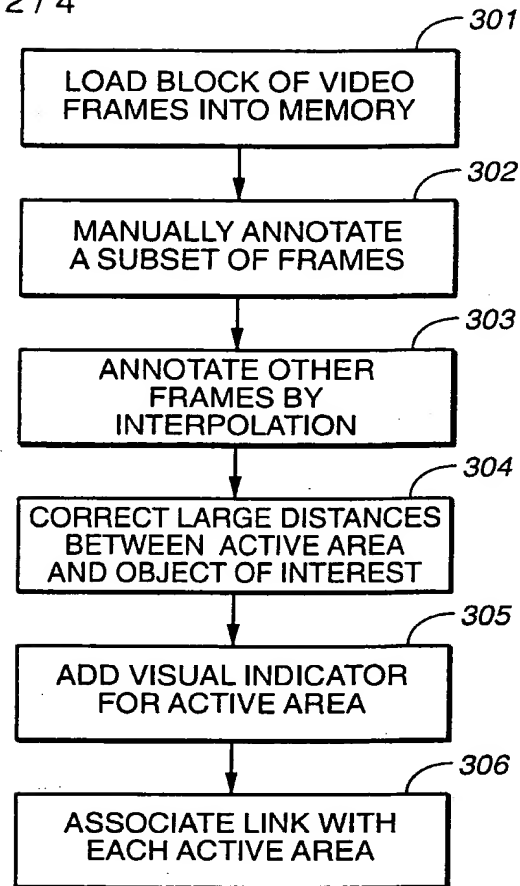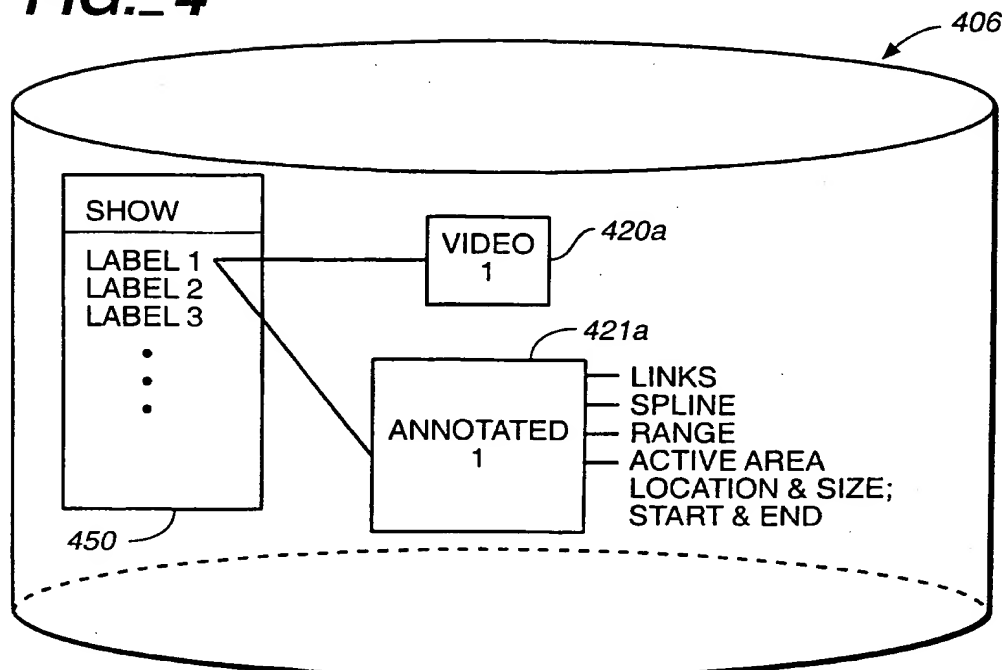22   location established during act (a) above.

23

24   44.   Apparatus of claim 40 wherein said memory storage device stores the location

25   of said active area for some of said frames but not others of said frames.

30

SUBSTITUTE SHEET (RULE 26)

FIG._1

FRAME

FIG._2

FIG._3

```
┌─────────────────────────────┐  ⌐ 301
│  LOAD BLOCK OF VIDEO        │
│  FRAMES INTO MEMORY         │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐  ⌐ 302
│  MANUALLY ANNOTATE          │
│  A SUBSET OF FRAMES         │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐  ⌐ 303
│  ANNOTATE OTHER             │
│  FRAMES BY                  │
│  INTERPOLATION              │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐  ⌐ 304
│  CORRECT LARGE DISTANCES    │
│  BETWEEN  ACTIVE AREA       │
│  AND OBJECT OF INTEREST     │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐  ⌐ 305
│  ADD VISUAL INDICATOR       │
│  FOR ACTIVE AREA            │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐  ⌐ 306
│  ASSOCIATE LINK WITH        │
│  EACH ACTIVE AREA           │
└─────────────────────────────┘
```

FIG._4

SHOW

LABEL 1
LABEL 2
LABEL 3
 ⋮

450

VIDEO 1 — 420a

ANNOTATED 1 — 421a

— LINKS
— SPLINE
— RANGE
— ACTIVE AREA
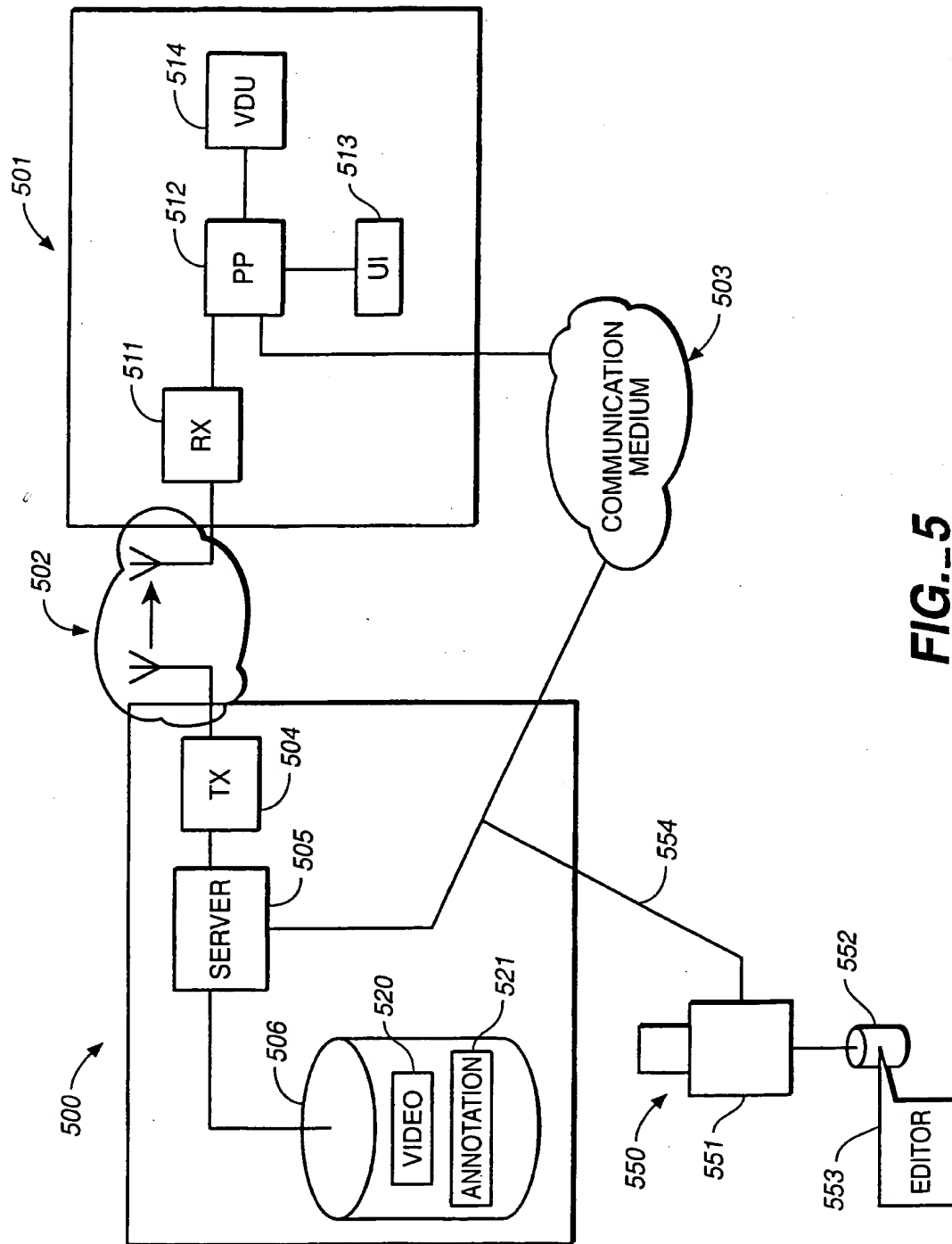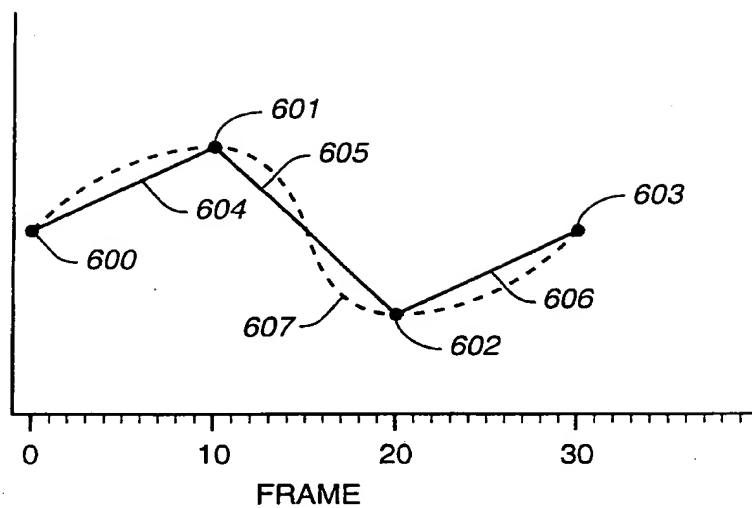 LOCATION & SIZE;
  START & END

406

FIG._5

FIG._6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/24263

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G06F 17/00
US CL :345/473, 474

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 345/473, 474, 418, 427

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US 6,006,241 A (PURNAVEJA et al) 21 December 1999, col.4, line 9 through col.9.	1-44
A	US 5,675,752 A (SCOTT et al) 07 October 1997.	1-44
A,P	US 6,032,156 A (MARCUS) 29 February 2000.	1-44

☐ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

*	Special categories of cited documents:
"A"	document defining the general state of the art which is not considered to be of particular relevance
"E"	earlier document published on or after the international filing date
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O"	document referring to an oral disclosure, use, exhibition or other means
"P"	document published prior to the international filing date but later than the priority date claimed

"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 DECEMBER 2000	09 JAN 2001

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	CLIFF NGUYEN VO
Facsimile No. (703) 305-3230	Telephone No. (703) 305-9594

Form PCT/ISA/210 (second sheet) (July 1998)